# Sound source localization identification accuracy: Envelope dependencies

William A. Yost[a)]
*Speech and Hearing Science, Arizona State University, P.O. Box 870102, Tempe, Arizona 85287, USA*

(Received 13 December 2016; revised 12 June 2017; accepted 15 June 2017; published online 13 July 2017)

Sound source localization accuracy as measured in an identification procedure in a front azimuth sound field was studied for click trains, modulated noises, and a modulated tonal carrier. Sound source localization accuracy was determined as a function of the number of clicks in a 64 Hz click train and click rate for a 500 ms duration click train. The clicks were either broadband or high-pass filtered. Sound source localization accuracy was also measured for a single broadband filtered click and compared to a similar broadband filtered, short-duration noise. Sound source localization accuracy was determined as a function of sinusoidal amplitude modulation and the "transposed" process of modulation of filtered noises and a 4 kHz tone. Different rates (16 to 512 Hz) of modulation (including unmodulated conditions) were used. Providing modulation for filtered click stimuli, filtered noises, and the 4 kHz tone had, at most, a very small effect on sound source localization accuracy. These data suggest that amplitude modulation, while providing information about interaural time differences in headphone studies, does not have much influence on sound source localization accuracy in a sound field. 
[http://dx.doi.org/10.1121/1.4990656]

[JFL] Pages: 173–185

## I. INTRODUCTION

Since 1971 (Yost *et al.*, 1971) a relatively large amount of literature has shown that the interaural time difference (ITD) of the envelope of carrier sounds [usually high-frequency (HF) carrier sounds where ITD fine-structure processing does not occur] influences discrimination of interaural differences and the laterality of the sounds when presented over headphones. These studies have been reviewed by many authors including Yost and Hafter (1987), Blauert (1997), and more recently by several other authors (e.g., Diedesch and Stecker, 2015; Bibee and Stecker, 2016), who have continued to study ITD envelope processing. ITD envelope processing led to a modification of the Duplex Theory of sound source localization that originally implicated ITD processing for low-frequency (LF) sounds and interaural level difference (ILD) processing for HF sounds. The idea of the Duplex Theory was originally proposed by Lord Rayleigh (1907) and reiterated by Stevens and Newman (1934). A modified version of the Duplex Theory of sound source localization states, "Interaural level difference is the cue used for locating HF sounds. ITD is the cue used to locate any sound with low frequencies or any HF complex sound with LF repetition in the time-domain waveform" (Yost, 2007, p. 183).

Two types of stimuli have been primarily used to implicate envelope ITD processing: click trains and amplitude modulated noises and tones. The earlier studies often used high-pass filtered clicks. The repetition rate and number of clicks in click trains have been studied. ITD and ILD thresholds and the laterality of the click trains have been investigated. As reviewed by many authors including Yost and Hafter (1987), Blauert (1997), and more recently Stecker and Bibee (2014), most of the data from these studies are consistent with temporal integration and adaptation processes often emphasizing early occurring clicks in a click train. The details of such adaptation processes are still being investigated (e.g., see Stecker and Bibee, 2014; Bibee and Stecker, 2016).

The studies of Bernstein and Trahiotis (e.g., see Bernstein and Trahiotis, 2014 for their most recent publication related to envelope ITD processing) have led the way in describing how amplitude modulation affects envelope ITD processing in lateralization studies. When the rate of the amplitude modulation of carriers (noise or HF tones) is less than approximately 300 Hz (but see Monaghan *et al.*, 2016), interaural difference thresholds and the laterality of the sounds can be based on envelope ITDs. The "shape" of the envelope has been shown (e.g., see Bernstein and Trahiotis, 2009) to be an important variable in determining envelope ITD processing as demonstrated in interaural discrimination and laterality experiments. The binaural system appears to be especially sensitive to the envelope of "transposed" processed stimuli, a particular modulation process introduced by Bernstein and Trahiotis (2002).

The vast majority of the literature on envelope ITD processing has involved stimuli presented over headphones. In these studies the ILDs of the stimuli have almost always be set to zero and the waveforms high-pass filtered in one way or another to reduce, if not eliminate, the role of temporal fine structure as a cue for ITD processing. For high-pass stimuli with ILDs set to zero, ITD discrimination thresholds are higher than when the stimuli are not high-pass filtered (i.e., when temporal fine structure cues can be used for ITD

[a)]Electronic mail: william.yost@asu.edu

processing), and laterality judgments can be difficult in these conditions. Providing envelope variables to these high-pass stimuli that have no ILDs often decreases ITD thresholds and makes laterality judgments possible. These changes in ITD discrimination and laterality performance as a function of different envelope variables is the basis for the implication that envelope ITDs can be a cue for binaural processing related to sound source localization in the front azimuth plane of a sound field. Functions relating envelope variables to ITD thresholds and laterality have been instrumental in refining models of ITD processing, especially cross-correlation models.

Very few studies (Eberle and Flanagan, 2000; Yost and Brown, 2013) have investigated the role of envelope variables on sound source localization in the front azimuth sound field. The stimuli used in both studies cited above were wideband noises. Since the noises were presented in a sound field, there were always ILDs (which can be large at high frequencies), and sound source localization accuracy is usually very good for broadband (BB) noise stimuli (e.g., see Stevens and Newman, 1934; Yost *et al.*, 2013). Most likely as a consequence of the presence of ILD cues and high sound source localization accuracy, neither paper (Eberle and Flanagan, 2000; Yost and Brown, 2013) found evidence for envelope ITD processing for sound source localization of a single BB noise in a sound field. However, the headphone studies that have implicated envelope ITD processing have used a variety of stimuli (including BB noises). Thus, in order to more fully evaluate the extent to which envelope ITDs may play a role in sound source localization accuracy measured in the front azimuth sound field, several experiments were conducted using stimuli like those used in the headphone studies (i.e., click trains; modulated, filtered noises, and modulated tones). It should be recognized that the main measure of performance in a sound field is sound source localization accuracy, and over headphones the measures are usually interaural discrimination thresholds and laterality judgments. Sound source location accuracy often cannot be measured for headphone-delivered stimuli, nor can interaural difference thresholds be measured in a sound field in the same way as the measures are made with headphone-delivered stimuli.

This paper is the most recent in a series of papers investigating the role of different stimulus parameters on sound source localization accuracy in the front azimuth sound field (Yost *et al.*, 2013; Yost and Zhong, 2014; Yost, 2016). The previous work suggests that stimulus bandwidth is a crucial variable affecting sound source localization. For narrow bandwidths, and only for narrow bandwidths (<2 octaves), the work suggests that center frequency (CF) (or tonal frequency) has a significant effect on sound source localization accuracy (Yost and Zhong, 2104). For bandwidths of two or more octaves the work indicates that sound source localization accuracy does not appear to depend on the CF of the noise (Yost *et al.*, 2013; Yost and Zhong, 2014). In this work neither duration nor overall level appear to have much of an effect on sound source localization accuracy in the azimuth plane independent of the sound's frequency content (Yost, 2016). The current paper concerns the role of amplitude modulation (envelope) on sound source localization accuracy. Given the results of the previous sound source localization accuracy papers and the headphone studies of envelope ITD processing, the bandwidth and frequency content of the stimuli and their interaction with envelope variables will be investigated, as bandwidth and frequency content (at narrow bandwidths) appear to be important variables affecting sound source localization accuracy (Yost and Zhong, 2014). With the exception of varying the number of clicks in a click train, stimulus duration and overall level will not be systematically manipulated in the studies reported in this paper as they do not appear to be important variables affecting sound source localization accuracy (Yost, 2016) in the front azimuth sound field.

Envelope and modulation in this paper will refer to the ongoing changes in amplitude during the time the sound is being presented. These terms are not intended in this paper to refer to the way in which the stimuli were turned on and/or off. In most conditions there was an onset and offset time difference between the ears due to the time of arrival difference when a sound source is at an azimuth angle relative to the listeners' head other than 0° (i.e., according to data like those from Kuhn, 1987, these time differences would be less than approximately 850 $\mu$s). For noise and tonal sounds, the stimuli were gated on and off with 20 ms rise/fall times which would reduce the role of the small ITDs (<850 $\mu$s) associated with sounds arriving at the ears from different azimuth locations (see Rakerd and Hartmann, 1986, who showed that rise times do affect ITD processing).

## II. EXPERIMENT I—CLICKS AND CLICK TRAINS

In experiment I, 100 $\mu$s clicks were filtered and presented as a single click or in click trains when the number of clicks and the rate of the clicks in the click train were varied. Two filter conditions were used: BB filtering (125 Hz to 8 kHz) and a 2 to 8 kHz HF, bandpass filtered (with a CF of 4 kHz) condition. Sound source localization accuracy using the procedure described in Yost *et al.* (2013) was used for a single click, for the number of clicks (2–16 clicks) in a click train presented at a click rate of 64 Hz, and for a 500 ms duration click train with clicks presented at rates of 16 to 128 Hz.

### A. Method

#### *1. Listeners*

In the test of a single click, 12 listeners were used (see Yost, 2016).[1] Nine listeners were females, and the ages of all listeners were between 19 and 32 yrs. An additional 12 listeners participated in the other two conditions of experiment I. There were eight females, and the ages of all listeners were between 19 and 38. All listeners reported that they had normal hearing. No listener in any experiment had previously participated in a hearing experiment. As explained in Secs. II A, III A, and IV A of the three experiments all listeners were provided a short training session. All of the conditions used in experiments I–III were approved by the

Arizona State University Institutional Board for the Protection of Human Subjects.

### 2. Stimuli

In the single-click condition (part of the Yost, 2016, study[1]) a single 100-$\mu$s click filtered from 125 Hz to 8 kHz (three-pole Butterworth Filter) was tested, and the results compared to data involving a 25 ms noise burst filtered the same as the BB clicks (see Yost, 2016). The level of an individual click for each filter condition was determined by matching (using an oscilloscope) the peak level of the click to the peak level of a 1000 Hz tone presented at 65 dB sound pressure level (SPL). Each single click, therefore, had a peak level equivalent to that of a 1000 Hz tone presented at 65 dB SPL. The level of the 25 ms noise burst was 65 dBA. All clicks in the click trains were 100 $\mu$s clicks, filtered with a three-pole Butterworth filter implemented in MATLAB (Mathworks, Natick, MA) either between 125 Hz and 8 kHz (BB) or between 2 and 8 kHz (HF). These filter conditions are the same used in Yost *et al*. (2013), Yost and Zhong (2014), and Yost (2016).

For the click trains the overall level of the click train was maintained at 40 dBA independent of the number of clicks (number of clicks greater than one) and click rate. The level measurements were made at the position of the listeners' heads in the listening room (see below). The two bandpass filtered clicks (BB and HF) were presented in click trains at 64 Hz (64 clicks/s, with a period of 15.625 ms) with 2 clicks (a 15.625 ms first click onset to last click onset time difference[2]), 4 clicks (46.875 ms first click onset to last click onset time difference[2]), 8 clicks (109.375 ms first click onset to last click onset time difference[2]), and 16 clicks (234.375 ms first click onset to last click onset time difference[2]). In the final condition, a 500 ms duration click train was used with click rates of 16 Hz (62.5 ms period, or 8 clicks in 500 ms), 32 Hz (31.25 ms period or 16 clicks in 500 ms), 64 Hz (15.625 ms period or 32 clicks in 500 ms), or 128 Hz (7.8125 ms period or 64 clicks in 500 ms).

### 3. Listening room

The same listening room used in Yost *et al*. (2015) was used in the present experiments. The room is a $10' \times 15' \times 10'$ lined on all six surfaces with acoustic foam (NRC rating of 0.96 and an absorption coefficient at 4 kHz of 1.44). The wideband reverberation time ($RT_{60}$) is 102 ms and $RT_{60}$ for a one-third octave noise centered at 4 kHz was 72 ms (BB and 4 kHz frequency regions were the primary frequency regions used in the current study). Twenty-four loudspeakers (Boston Acoustics 100×, Peabody, MA) are on a 5 ft radius circle (i.e., azimuth array with 15° loudspeaker spacing) at the height of listeners' pinnae, but only the 13 loudspeakers in the front azimuth sound field were used in this experiment. There is a control room from which listeners are monitored by an intercom and camera. Listeners were instructed at all times to face the center loudspeaker that had a red circle on its center. They were monitored on each trial and rarely failed to face the center loudspeaker. All sounds were presented via a 24 channel Digital-to-Analog (DA) converter (two, Echo Gina 12 DAs, Santa Barbara, CA) outputting sounds at a rate of 44 100 samples/s/channel.

### 4. Procedure

The sound source localization accuracy procedure of Yost *et al*. (2013) was used. A sound was presented from one of 11 loudspeakers (separated by 15° between ±75°) in the front azimuth sound field (see Fig. 1). Listeners were instructed to indicate the location (via a keyboard) of the loudspeaker (loudspeakers, #1 to #13, had numbers on them) that presented the sound. They were told that sound could come from loudspeakers #1 (90°, far right) to #13 (−90°, far left), but they were not aware that sound was presented only from loudspeakers #2 to #12 (±75°). This process was to protect against "edge-effects" (see Hartmann *et al*., 1998).

In the single-click condition (part of the Yost, 2016, study[1]), a single click was presented on each trial randomly intermixed with noises of different durations (25 to 450 ms; see Yost, 2016).[1] Each stimulus/loudspeaker combination appeared 20 times in random order. In the condition where the number of clicks was varied, the five different number of click cases were randomly presented 15 times for each of the 11 loudspeaker locations (165 trials per number of clicks or 825 trials/listener) chosen randomly. These 825 trials were divided into 11, 75-trial blocks. The condition in which click rate was varied was similar to the number-of-click conditions in that the four click rates were each presented 15 times for each of the randomly chosen 11 loudspeaker locations (660 trials/listener, divided into ten, 66-trial blocks). All listeners started their participation in the experiments with a practice session in which a 200 ms BB noise was presented in succession from each of the 13 loudspeakers (#1 far right to #13 far left, see Fig. 1). They were able to listen to these 13 noise bursts as many times as they wished to acquaint them with the sounds and their locations.

## B. Results

The main part of Fig. 2 displays data for the number-of-click conditions indicating mean (12 listeners) root-mean-squared (rms) error (degs) and one standard deviation of rms error (degs) as a function of the number of clicks presented

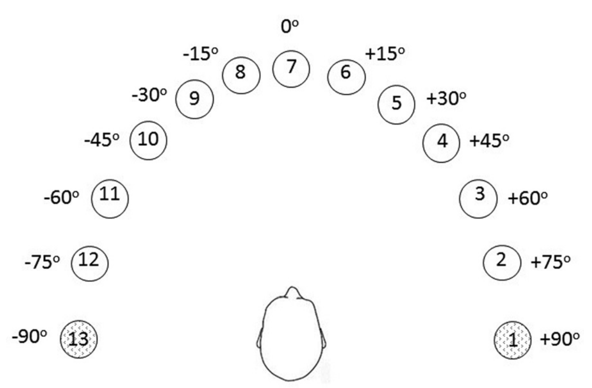


FIG. 1. Diagram of the 13 loudspeakers in the front azimuth field located at pinna height. Loudspeakers spaced 15° apart from 90° (loudspeaker #1) to −90° (#13). Loudspeakers #1 and #13 did not present sound. Listeners did not know this and could indicate that a sound came from these loudspeakers.

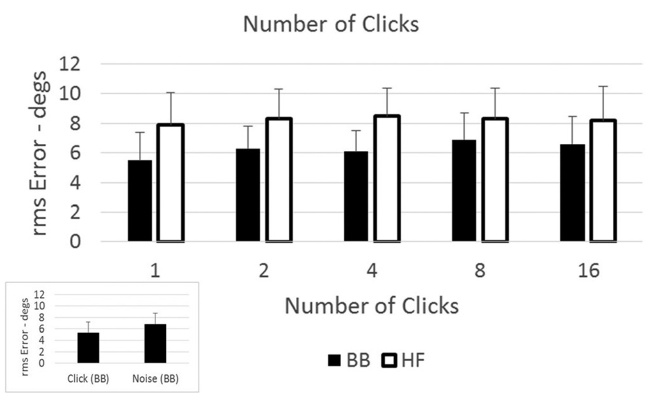


FIG. 2. Mean (12 listeners) rms error (degs) as a function of number of clicks and filtering condition (BB; and HF, bandpass filtered). The inset are data from Yost, (2016)[1] indicating mean rms error (degs) for a BB click and a BB noise. Error bars are one standard deviation (calculated across listeners).

at a rate of 64 Hz. Data from the BB and HF frequency filter conditions are shown. The inset includes data for the single-click condition (from Yost, 2016)[1] comparing rms error (degs) for a single click as compared to a 25 ms noise burst. Both stimuli for the data in the inset were filtered between 125 Hz and 8 kHz (BB). The data in the inset represent the results from 12 listeners (see Yost, 2016).[1] Sound source localization accuracy as measured by rms error shown in the figures of the current paper represent the rms error averaged over all 11 loudspeaker locations [see Fig. 1, and see Sec. V B, and point (1) for additional data related to rms errors as a function of loudspeaker location].

The data of Fig. 2 suggest that sound source localization accuracy depends on filtering the click, but that accuracy does not depend on the number of clicks. All 12 listeners who generated the data in the inset of Fig. 2 had rms errors that were smaller for the single click than for the 25 ms noise burst. All 12 listeners who participated in the number-of-clicks condition (Fig. 2) had lower rms errors for the BB as compared to the HF conditions. No listener showed either a consistent increase or decrease in the rms error as a function of the number of clicks for either filter condition.

Figure 3 displays data for the click-rate condition indicating mean (12 listeners) rms error (degs) and one standard deviation of rms error (degs) as a function of click rate for a 500 ms click train. Data from the BB and HF filter cases are

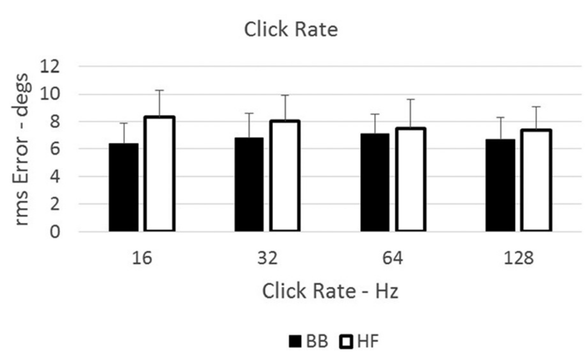


FIG. 3. Similar to Fig. 2 indicating mean (12 listeners) rms error (degs) as a function of click rate and filtering condition (BB; and HF, bandpass filtered). Error bars are one standard deviation (calculated across listeners).

shown. The data of Fig. 3 suggest that while filtering clicks does affect sound source localization accuracy, changing click rate does not at least with duration held constant. All 12 listeners who participated in the click-rate condition had lower rms errors for the BB as compared to the HF conditions except for the 128 Hz rate, in which case only eight of the 12 listeners had lower rms errors for the BB as compared to the HF condition. No listener showed either a consistent increase or decrease in rms error as a function of click rate for either filter condition.

A two-way (filter condition and number of clicks) repeated measures Analysis of Variance (ANOVA) was conducted for the data of Fig. 2 with a 0.05 level of significance. There was a significant statistical difference based on filter condition, but not on the number of clicks. The interaction between filter condition and number of clicks was not statistically significant. For the data from Yost (2016, the inset in Fig. 2) the rms error for the single click was significantly less than that for the 25 ms noise burst using a repeated measures $t$-test (at the 0.05 level of significance). The rms error for localizing the click stimulus was always less than or equal to that for the noise stimulus for all 12 listeners in the Yost (2016) study. A between-subjects $t$-test indicated that the rms error for the single click condition (inset to Fig. 2, rms error of 5.6°) measured in Yost (2016) was not statistically different from that measured in the current study (main Fig. 2, rms error of 5.7°). Another two-way (filter condition by click rate) ANOVA was conducted with a 0.05 level of significance for the third condition (see Fig. 3). The filter condition yielded a statistically significant change in rms error, but not click. There was no statistically significant interaction.

## III. EXPERIMENT II—AMPLITUDE MODULATED NOISE

In experiment II, sinusoidal and transposed (see Bernstein and Trahiotis, 2002) amplitude modulation procedures for filtered noises were used in the sound source localization accuracy procedure of Yost *et al.* (2013). Sinusoidal and transposed modulation processes were also used for a BB noise (125 Hz–8 kHz). The transposed modulated stimulus process was applied to two-octave wide and 1/10-octave wide 500 ms noise bursts with CFs of 250 Hz (LF noise) and 4 kHz (HF noise). In all cases unmodulated (UM) noises were also tested. Again, these filter conditions are the same used by Yost and Zhong (2014) and Yost (2016).

### A. Methods

#### 1. Listeners

Twelve listeners were used for Experiment II, eight females and four males all between the ages of 20 and 39. No listener in Experiment II participated in Experiment I. All listeners reported having normal hearing and had no prior experience in hearing experiments.

#### 2. Stimuli

Independently generated 500 ms noise bursts were used. All of the noises were modulated with the transposed procedure as described by Bernstein and Trahiotis (2002). The transposed procedure involves linear half-wave rectification of the time-domain waveform of a LF tone with a frequency equal to the rate of modulation (16, 32, 64, 128, 256, and 512 Hz). The rectified waveform was then transformed to the frequency domain, and the magnitudes of components above 2 kHz were filtered out by setting them to zero. The resulting filtered signal was transformed back to the time domain and multiplied by a filtered noise carrier. These transposed processed stimuli have an envelope that is similar to that of the rectified and filtered pure tone (see Bernstein and Trahiotis, 2002, for additional details). The BB noise bursts were also sinusoidally amplitude modulated (SAM) at rates of 16, 32, 64, 128, 256, and 512 Hz, always with the modulating tone having a 0° starting phase, and the overall level was kept constant (modified by the modulation rate). UM filtered noises were also used. All filtered and modulated noise bursts were further filtered with three-pole Butterworth filters as follows: a 1/10 octave wide filter centered at 250 Hz (1/10, 250), a 1/10 octave wide filter centered at 4 kHz (1/10, 4 k), and a BB filter (125 to 8 kHz, BB). Thus, there were four stimulus conditions in experiment II with the 1/10, 250; 1/10, 4 k; and BB stimuli modulated with the transposed envelope (BB-Trans.) and the BB stimulus also modulated with a sine tone (BB-SAM). Rates of modulated were UM, 16, 32, 64, 128, 256, and 512 Hz. All stimuli were presented with 20 ms cosine-squared, rise-fall times at 40 dBA. The depth of modulation in all cases was 100%.

### 3. Listening room

The same listening room was used as was used in experiment I.

### 4. Procedure

The same procedure and practice session described for experiment I were used to present noise bursts to the different loudspeakers in experiment II. There were 28 stimulus conditions in experiment II [four filter/modulation conditions (1/10 oct-250 Hz, 1/10 oct-4 kHz, BB-Transposed, BB-SAM) by seven modulation rates (UM, 16, 32, 64, 128, 256, 512 Hz)]. These 28 conditions were presented in random order 15 times each to the 11 loudspeakers chosen at random for 4620 trials/listener. These trials were divided into sixty, 77 trial blocks tested over several sessions across several days depending on listeners' schedules. As in experiment I, listeners indicated which of 13 loudspeakers (see Fig. 1) may have presented a noise burst.

## B. Results

Figure 4 displays data for experiment II indicating mean (12 listeners) rms error (degs) and one standard deviation of rms error (degs) as a function of the modulation rate (UM, 16, 32, 64, 128, 256, and 512 Hz) and filter condition (1/10, 250; 1/10, 4 k; and BB for the transposed processed envelope condition). The data of Fig. 4 suggest that sound source localization accuracy depends on filtering the noise as is consistent with Yost and Zhong (2014), but that accuracy may

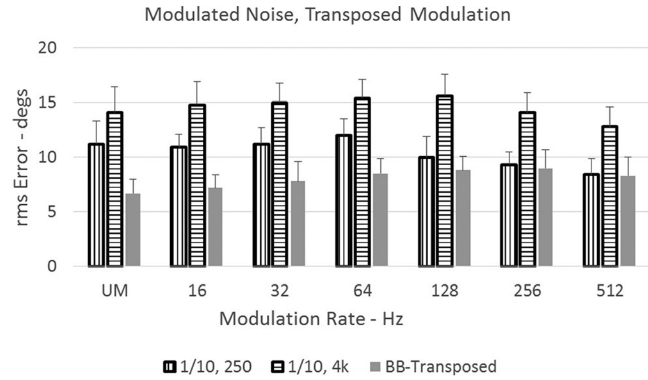


FIG. 4. Mean (12 listeners) rms error (degs) as a function of modulation rate (UM) and filtering condition (1/10 octave filter with 250 Hz CF, 1/10 octave filter with 4 kHz CF, and BB clicks). All stimuli were modulated with the transposed method. Error bars are one standard deviation (calculated across listeners).

not depend on the modulation rate. All 12 listeners when the modulation rate was less than 128 Hz had the lowest rms errors for the BB-Transposed condition and second lowest rms errors for the 1/10, 250 condition. At 128 to 512 Hz modulation rate the 1/10, 4 k condition had the highest rms error for all 12 listeners. The number of listeners whose rms errors were higher for the 1/10, 250 Hz condition as compared to the BB-Transposed condition were: eight listeners at 128 Hz, six listeners at 256 Hz, and five listeners at 512 Hz.

Figure 5 displays data for experiment II indicating mean (12 listeners) rms error (degs) and one standard deviation of rms error (degs) as a function of the modulation rate (UM, 16, 32, 64, 128, and 256 Hz) comparing the BB stimulus modulated with the transposed method (BB-Trans.) and a sinusoidally modulated (BB-SAM) envelope. The "BB-Transposed" data in Fig. 4 are the same data shown in Fig. 5 as the "BB-Trans." data. The data of Fig. 5 suggest that sound source localization accuracy does not depend on the form of modulation as least in terms of using the transposed method or SAM modulation of a BB noise source. The rms errors shown for the UM conditions in Figs. 4 and 5 are very similar to those measured by Yost and Zhong (2014) for the same stimulus conditions. None of the 12 listeners had a

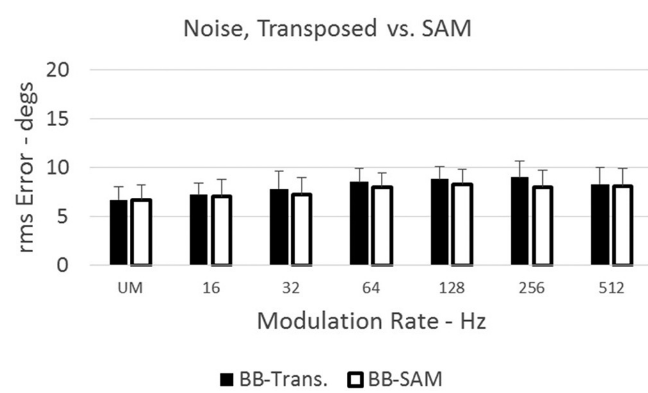


FIG. 5. Mean (12 listeners) rms error (degs) as a function of modulation rate (UM) and type of modulation (Transposed, Trans., or SAM). Stimuli were BB. Error bars are one standard deviation (calculated across listeners).

consistent increase or decrease in rms error as a function of the modulation rate for either type of modulation.

A two-way (filter condition and modulation rate) repeated measures ANOVA was conducted for the data of Fig. 4 with a 0.05 level of significance. There was a significant statistical difference based on filter condition and modulation rate. The interaction between filter condition and modulation rate was statistically significant. The mean rms errors across modulation rates were compared in pair-wise *t*-tests for the three filter conditions (1/10, 250; 1/10, 4 k; and BB filter conditions). The differences for all three comparisons were statistically significant (i.e., the 1/10, 4 k case produced the largest errors, followed by the 1/10, 250 case, and then the BB case). The statistically significant difference due to the modulation rate is most likely due to the decrease in rms error as the modulation rate increased from 128 to 512 Hz, especially for the 1/10, 4 k condition (which probably partially accounts for the statistically significant interaction).

A two-way (modulation type and modulation rate) repeated measures ANOVA was conducted for the data of Fig. 5 with a 0.05 level of significance. There was not a significant statistical difference based on the type of modulation (SAM or Transposed) nor was there a statistically significant difference based on the modulation rate. The interaction between filter condition and modulation rate was not statistically significant.

## IV. EXPERIMENT III—AMPLITUDE MODULATED TONE

In experiment III sound source localization accuracy (Yost *et al.*, 2013) was measured using the transposed procedure (see experiment II) for a 4 kHz carrier tone and an UM 4 kHz tone. Modulation rates of 16, 32, 64, 128, and 256 Hz were tested.

### A. Methods

#### *1. Listeners*

Twelve listeners were used for experiment III; nine females and three males all between the ages of 19 and 28. All listeners reported having normal hearing, and no listener had previously participated in a hearing experiment. All listeners in experiment III were different from those used in experiments I and II.

#### *2. Stimuli*

The same transposed procedure described for experiment II was used to modulate a 500 ms, 4 kHz carrier tone at rates of 16, 32, 64, 128, and 256 Hz. The 4 kHz carrier tone was presented with zero degrees starting phase. An UM 4 kHz tone was also used. The overall level of the sound was 40 dBA for all modulation rates and all sounds were gated with a 20 ms cosine squared rise-fall time. The depth of modulation was always 100%.

#### *3. Listening room*

The same room as was used in experiments I and II was used in experiment III.

#### *4. Procedure*

The same procedure and practice session described for experiments I and II were used to present the modulated and UM 4 kHz tone to the different loudspeakers in experiment III. The six randomized modulation rate (UM, 16, 32, 64, 128, and 256 Hz) conditions were presented 15 times for each of the 11 randomly chosen loudspeakers yielding 990 trials/listener, divided into 11, 90-trial blocks. As in experiments I and II, listeners indicated which of 13 loudspeakers (see Fig. 1) may have presented a tone.

### B. Results

Figure 6 displays data for experiment III indicating mean (12 listeners) rms error (degs) and one standard deviation of rms error (degs) as a function of the modulation rate (UM, 16, 32, 64, 128, and 256 Hz). The data of Fig. 6 suggest that sound source localization accuracy may be slightly better when the transposed method was used for the 4 kHz tone than when the tone was UM (e.g., the 16 Hz modulated tone has 1.2° less rms error than the UM condition). It appears as if sound source localization accuracy decreases slightly with increasing modulation rate (e.g., a decrease in mean rms error of 2.8° from 16 to 256 Hz).

The 16 Hz transposed processed 4 kHz stimulus had a lower rms error than the UM tone for eight of the 12 listeners. Five of these eight listeners showed the same trend (i.e., decreasing rms error with increasing modulation rate) in rms error from 64 to 256 Hz as shown for the mean data in Fig. 6. The UM rms error shown in Fig. 6 is very similar to that obtained by Yost and Zhong (2014) at 4 kHz.

A one-way (modulation rate) repeated measures ANOVA was conducted for the data of Fig. 6 with a 0.05 level of significance. There was a significant statistical difference based on the modulation rate. There was a statistically significant difference in rms error between the UM and 16 Hz modulation cases, and a statistically significant difference between the 16 Hz modulation case and the 256 Hz

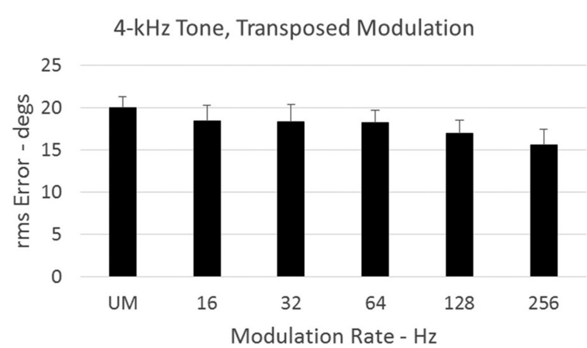


FIG. 6. Mean (12 listeners) rms error (degs) as a function of modulation rate (UM) for a transposed modulated 4 kHz tonal carrier. Error bars are one standard deviation (calculated across listeners).

modulation case. These two statistical differences were based on *post hoc* *t*-tests.

## V. OVERALL DISCUSSION

Sound source localization accuracy for click trains does not appear to vary as a function of the number of clicks (Fig. 2) nor as a function of click rate (Fig. 3), over the range of conditions used in the present study. The lack of an effect of number of clicks or click rate on accuracy does not depend on the filtering condition. Measures of ITD discrimination thresholds over headphones do change as a function of number of high-pass filtered clicks. ITD discrimination thresholds measured with headphone-delivered stimuli decrease as the number of HF, bandpass clicks increase (from 1 to 64) when the click rate is between 100 and 1000 Hz, (e.g., Hafter *et al.*, 1983; Yost, 1976). And, ITD discrimination thresholds increase as the HF, bandpass click rate increases (click rates from 50 to 450 Hz in Yost, 1976, 100 to 1000 Hz in Hafter *et al.*, 1983). There is an interaction between the number of HF, bandpass clicks and click rate (e.g., see Hafter *et al.*, 1983), i.e., the effect of HF, bandpass click rate on ITD discrimination thresholds increases with increasing number of clicks (from 2 to 32 clicks).

Sound source localization accuracy for locating filtered noise bursts does not appear to depend on either the transposed or SAM modulation methods for modulation rates ranging from 16 to 512 Hz independent of the filter conditions (Figs. 4 and 5). That is, even for the narrowband 1/10 octave wide noise bursts when sound source localization accuracy is poor for UM noises, modulation does not improve sound source localization accuracy above that which is probably provided by increased bandwidth when the modulation rate increases. However, ITD discrimination thresholds measured over headphones are lower for modulated than for UM HF noise bursts, and such thresholds can vary as a function of the modulation rate, e.g., envelope ITD thresholds increase with increases in the modulation rate above 128 Hz (e.g., Bernstein and Trahiotis, 1994). Envelope ITD discrimination thresholds are generally lower when the transposed method is used as opposed to the SAM modulation method (Bernstein and Trahiotis, 2002).

Sound source localization accuracy did vary as a function of using the transposed modulation method for a 4 kHz tonal carrier stimulus (Fig. 6). There was a statistically significant small 1.2° decrease in rms error for the 16 Hz modulated tone as compared to the UM tone. Accuracy increased (rms error decreased) by an average of 2.8° as the modulation rate increased from 16 to 256 Hz. Envelope ITD discrimination thresholds measured over headphones cannot be measured for an UM 4 kHz tone and are as low as approximately 50 $\mu$s for modulated tones, but are most often 100 $\mu$s or more (Bernstein and Trahiotis, 2002). Envelope ITD discrimination thresholds often decrease as the modulation rate increases from 16 to 128 Hz, and then increase (often rapidly) as the modulation rate continues to increase from 128 to 512 Hz (e.g., see Bernstein and Trahiotis, 1994, but also see Monaghan *et al.*, 2016).

Thus, sound source localization accuracy does not depend on the characteristics of modulation in the same way as envelope ITD discrimination thresholds depend on these same modulation characteristics when the sounds are presented over headphones. In general, sound source localization accuracy varies very little, or does not vary at all (e.g., for some listeners), when carrier sounds are modulated including presenting clicks in click trains. The small changes in sound source localization accuracy that do occur when there is a modulated envelope do not usually follow the same pattern of change as when envelope ITD discrimination thresholds are measured for the same envelope characteristics (e.g., modulation rate). As documented above, envelope ITD discrimination thresholds often increase with increasing modulation rate (especially as modulation rates increase above 128 Hz), but rms errors for sound source localization accuracy sometimes decrease with increasing modulation rate (Figs. 4 and 6). Or, ITD discrimination thresholds increase with an increasing high-pass filtered click rate, but sound source localization accuracy does not appear to depend on click rate (Fig. 3).

Several studies (e.g., see Trahiotis and Stern, 1989; Bernstein and Trahiotis, 1985; Buell *et al.*, 1994) have investigated the laterality of modulated stimuli presented over headphones. That is, listeners have been asked to indicate the lateral (within the head) locations of sounds presented with envelope ITD differences. In general, the results of these laterality experiments are consistent with those from studies of envelope ITD discrimination thresholds, indicating that listeners can determine changes in the lateral position of images within the head consistent with how envelope ITD discrimination thresholds vary (Bernstein and Trahiotis, 1985).

### A. Differences between sound field and headphone-delivered psychophysical measures

There are several possible differences between the current sound field study and those studies using headphone delivered stimuli that could affect the results of the current study:

(1) Probably the most important difference is that in a sound field there is always a non-zero (and often quite large) ILD, whereas in the vast majority of the headphone studies the ILD was set to zero. Thus, it appears as if providing an envelope in addition to the availability of an ILD cue does not have much of an effect on sound source localization accuracy in a sound field.

(2) For single BB sounds (e.g., a single BB, noise burst, or a single BB click), it might be that sound source localization accuracy performance in a sound field is already so good (at ceiling) without modulation that providing modulation cannot lead to any improvements [but see points (8) and (9) below].

(3) Providing a possible envelope ITD cue for sounds presented in a sound field in order to affect sound source localization accuracy may also be limited by the magnitude of the effects measured over headphones. In many of the headphone studies envelope ITD thresholds can be

more than several hundred microseconds. According to measurements like those made by Kuhn (1987), the time it takes sound to travel from one ear to the other can be as fast as 13 $\mu$s for HF sounds. Thus, it may be that some envelope ITDs that occur in a sound field are too fast to influence sound source localization accuracy due to a lack of sufficient interaural temporal acuity for envelope ITD processing.

(4) The loudspeakers and especially the room (reflective environment) can reduce the depth of modulation of any modulated sound (e.g., see Houtgast and Steeneken, 1973). Thus, poor sound source localization accuracy might be due to a lack of sufficient depth of modulation. The listening room used in the current experiments had low RTs especially at 4 kHz. Figure 7 displays one period of a 16 Hz modulated 4 kHz tonal stimulus using the transposed process at the input to the loudspeakers [Fig. 7(A)] and as measured in the room at the position of the listener's head [Fig. 7(B)] by a dual-ribbon microphone (Beyerdynamic M-160, Heilbronn Germany). The Room Response clearly shows the combined effect of the loudspeaker, room, and recording microphone on the shape of the envelope. But there is still a clear 100% depth of modulation. Thus, it is not likely that any diminution of the depth of amplitude modulation at 16 Hz had much of an effect on localization accuracy. As the modulation rate increases then the depth of modulation would probably decrease predicting an increase in rms error. But the results show an opposite trend, i.e., a decrease in rms error with increasing modulation rate (see Fig. 6). Thus, it is unlikely that the room (and/or the loudspeakers) had any substantial effect on sound source localization accuracy in the current study. This might not be true in a more reverberant space, assuming that amplitude modulation has an effect on sound source localization accuracy, which the other data of the current study suggest it does not. That is, in order for room acoustics to influence sound source localization accuracy based on the depth of sound's amplitude modulation, amplitude modulation must have an effect in a very dead room (anechoic) and then its effect diminished as the RT (or other measures of room acoustics) changes. The data of the present study cast some doubt as to how much of an effect a sound's modulation pattern has on sound source localization accuracy in a sound field even if the sound field is anechoic. A more definitive answer will require more study.

(5) Reflections in a room may make the ITDs incoherent to the extent that the incoherence reduces the ability to use ITDs in either the fine-structure or the envelope. However, the ILDs would be less affected by reflections (see Rakerd and Hartmann, 2010). In addition, measures of sound source localization accuracy in the same listening room for LF sounds where ITD fine-structure is the probable cause for sound source localization indicate approximately the same level of sound source

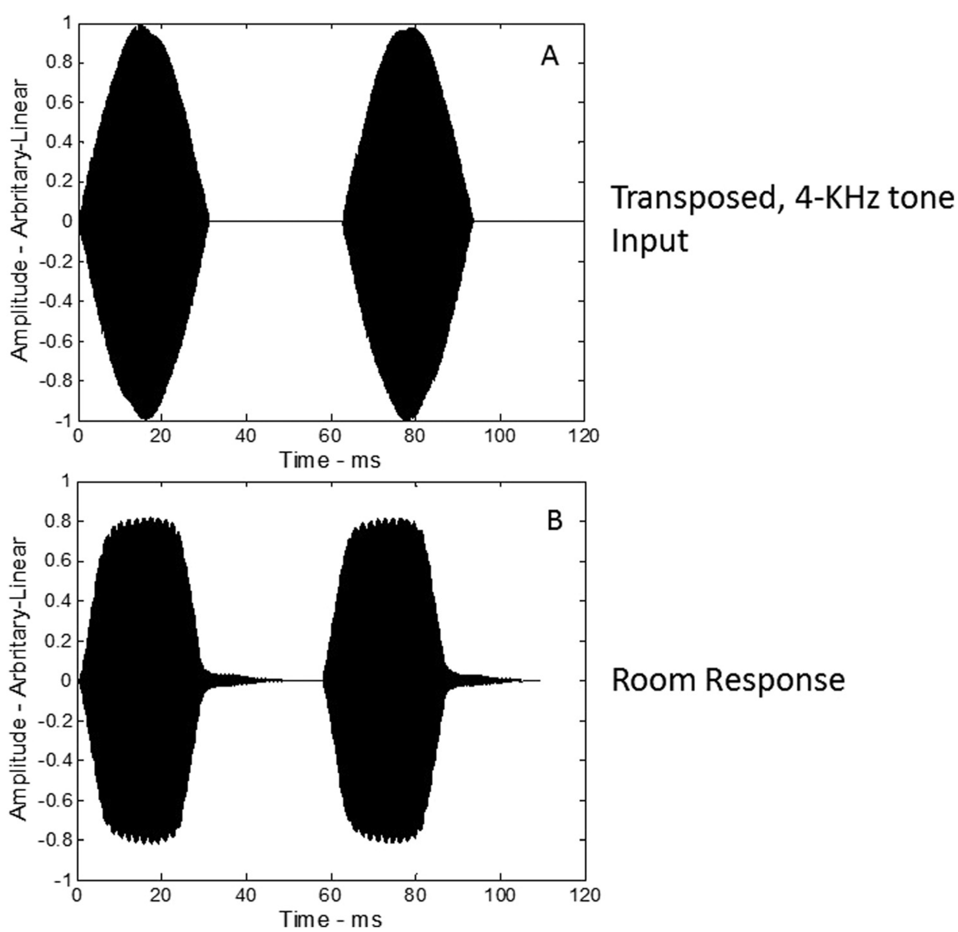


FIG. 7. (A) The transposed method was used to modulate a 4 kHz stimulus with a modulated rate of 16 Hz (62.5 ms period). (B) The measured modulated 4 kHz stimulus recorded at the listeners head. The waveform in (B) is based on the combined impulse response of the loudspeaker, listening room, and recording microphone (see text for details).

localization performance as measured in an anechoic space. For instance, Stevens and Newman's (1934) measurement in an echo-free, open space for a 200 Hz tone was about 11° of error and the Yost and Zhong (2014) measurement in the same room as used in the present paper was approximately 13° for a 250 Hz tone. Thus, whatever decorrelation there may be of ITD cues in the room used for the current study, the effects are probably very small and could be overcome by ILDs. Again, it seems unlikely that any such decorrelation would be differentially affected by waveform modulation.

(6) In the beginning of Sec. V it was mentioned that the decrease in rms error (degs) with increasing modulation rate (Figs. 4 and 6) may be due to increasing stimulus bandwidth, not modulation rate *per se*. Figure 8 illustrates this conjecture. Figure 8 shows the magnitude (dB) spectrum of a 4 kHz carrier tone modulated with the transposed method at 16 and 256 Hz rates of modulation. As can be seen, the sidebands for the 16 Hz modulation conditions are much reduced in magnitude and much closer in frequency to 4 kHz (the carrier) than those for the 256 Hz modulation rate. That is, the bandwidth of the stimuli is greater for 256 Hz modulation than for 16 Hz. Since wider bandwidths lead to lower sound source localization rms errors, the decrease in accuracy between 16 and 256 Hz of amplitude modulation could be due to the widening bandwidth. Yost and Zhong (2014) showed that sound source localization accuracy clearly depends on stimulus bandwidth for filtered noises and they discuss several reasons why this might be the case. Also note, modulation had no effect on sound source localization accuracy for the BB conditions. That is, when the stimulus is already BB making the spectrum broader by providing amplitude modulation has no effect on sound source localization accuracy.

(7) LF energy in the skirt of the low-pass filter (e.g., 2 kHz cutoff) used to filter the noise and click trains might allow for both ITD and ILD processing to affect sound source localization accuracy for HF stimuli. As a consequence there might be better accuracy in these cases in that energy at low frequencies might allow for processing ITD in the temporal fine structure of the stimuli. This is probably not relevant to measures of sound source localization accuracy in this paper. First, because frequency region, and therefore by implication the use of interaural cues (i.e., ITD at low frequencies and ILD at high frequencies), does not affect sound source localization accuracy for stimuli with bandwidths 2 octaves or wider as was the case in all but one condition in the present paper (Yost *et al.*, 2013; Yost and Zhong, 2014). For the 1/10 octave wide filter condition with a 4 kHz CF, sound source localization accuracy is worse than for wideband stimuli and for stimuli with the same bandwidth but with a 250 Hz CF (see also Yost and Zhong, 2014). However, the overall level of all stimuli in the present study was kept low (e.g., 40 dBA) meaning the spectrum level was very low. Filtering at approximately 18 dB/octave (three-pole filter) means the spectrum level of these sounds below the cutoff frequency of 2 kHz was very low (e.g., a 1/10 octave bandwidth at 4 kHz, 3.9 to 4.2 kHz, signal at an overall level of 40 dB SPL would have a spectrum level of approximately 15 dB SPL and at 1 kHz given the 2 kHz low-pass filter cutoff, the spectrum level would be approximately −3 dB SPL). Finally, for the transposed stimulus process the amplitudes of components of the spectrum of the rectified sine waves above 2 kHz were set to zero, narrowing the spectral bandwidth. While these low levels do not completely eliminate the possibility that accuracy performance could be affected by ITD at low-frequencies, it probably reduces this possibility a great deal.

(8) While sound source localization accuracy is very good for BB stimuli [see (2) above], such performance is poorer for narrowband noises (i.e., 1/10, 250 and 1/10, 4 k conditions), HF click trains, and the 4 kHz tone (see also Stevens and Newman, 1934; Blauert, 1997; Yost and Zhong, 2014). But the data of this study indicate that providing an envelope in these cases leads to at best a 1°–2° improvement in accuracy and only for some listeners and some conditions, when rms errors could on average be as much as 13° better (e.g., rms error for a 4 kHz UM tone was about 20° and the lowest rms error

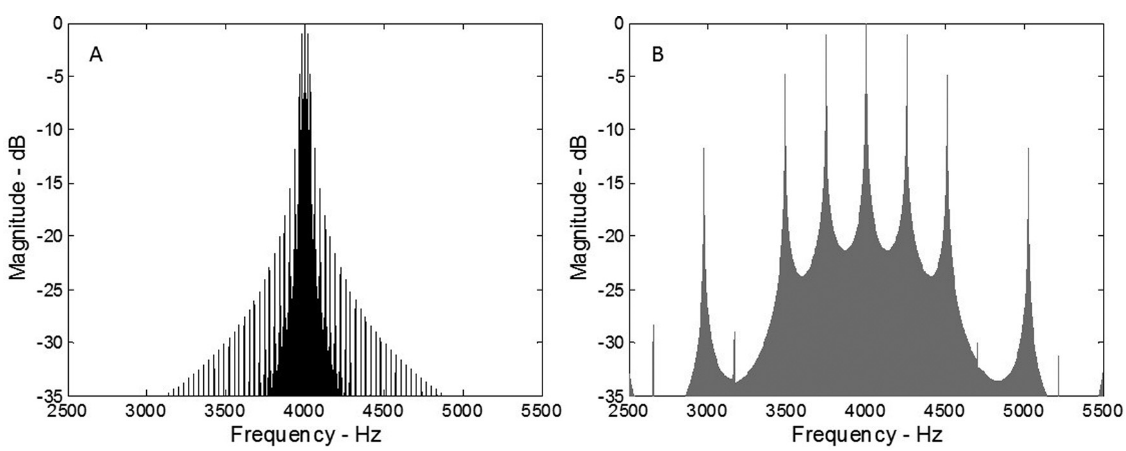


FIG. 8. Magnitude spectra (dB) using the transposed method for a modulated 4 kHz tonal carrier. (A) modulation rate is 16 Hz and (B) modulation rate is 256 Hz.

for a wideband noise is near 7°). The 1/10 octave wide stimuli were by virtue of narrowband filtering already amplitude modulated (see Rice, 1954) before a modulator was applied. It may be that the modulation provided by the modulator introduces very little additional modulation that could have increased sound source localization accuracy. However, recall that sound source localization for the 1/10 octave wide noises is poor compared to BB noise stimuli (e.g., Yost and Zhong, 2014).

(9) In the case of the HF clicks (2–8 kHz; see Figs. 2 and 3), sound source localization rms error was on average 7.9° while the average rms error for the BB clicks (125 Hz–8 kHz) was 6.5° (i.e., HF clicks produced approximately 1.4° greater rms error than BB clicks). Thus, it might have been possible for listeners to gain a slight increase in sound source localization accuracy (1.4° of rms error gain) through temporal integration of information as the number of clicks increased, but as shown in Fig. 2 listeners' performance did not improve. Listeners' interaural discrimination thresholds do improve with an increasing number of clicks in most headphone studies (see Yost and Hafter, 1987, for a review), but in these studies thresholds for a single click are almost always much higher (e.g., due to high-pass filtering the click and measuring ITD thresholds) than when the clicks are not high-pass filtered. Thus, there was "room" for improvement due to temporal integration caused by increasing the number of clicks in these headphone studies. There does not seem to be a clear reason for the lack of improved sound source localization accuracy in the current study using click trains, except that it might not be possible to measure a change in performance when performance is near ceiling and/or when there is only a 1.4° range over which performance can vary.

### B. Other aspects of the measurement of sound source localization accuracy in this study

(1) Listeners' heads were not constrained from moving in this study and neither was head motion recorded. Listeners were instructed to keep their heads fixed and always facing the front loudspeakers (on which was a large red circle). Listeners were monitored during the experiments and they rarely appeared to move their heads and were reminded not to do so when they did. With stimulus durations as long as 500 ms, there would have been time for the head to move toward the sides while a sound was on. Sound source localization accuracy is better when sound sources are in front of as opposed to the side of listeners (Stevens and Newman, 1934; Mills, 1958). Thus, accuracy performance may have been better than if the head had been restrained. However, the data in Fig. 9 suggests that such head movements, if they did play a role, probably played only a minor one. Figure 9 shows the mean rms error and one standard deviation of rms error as a function of the position of the loudspeaker (assuming the listener faced the

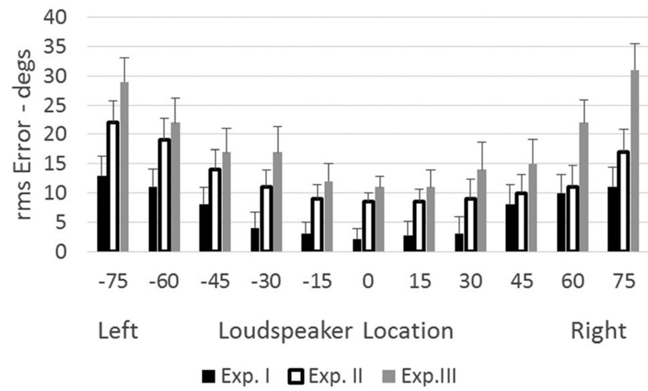


FIG. 9. Mean rms error (degs) as a function of loudspeaker location (see Fig. 1). Data are averaged across all conditions of experiment I (exp. I), across all conditions of experiment II (exp. II), and across all condition of experiment III (exp. III). Error bars are one standard deviation (calculated across listeners).

center loudspeaker at 0°, see Fig. 1, which they almost always did) for the average data of experiments I, II, and III. These data are consistent with other data (see Yost *et al.*, 2013 for a discussion of this literature) in the study of sound source localization accuracy in the front azimuth field indicating higher acuity (low rms error) for sounds presented from in front (near 0°) as opposed to off to one side (near ±75°). These results are consistent with the observations and arguments made by Mills (1958) and Stevens and Newman (1934) concerning sound source localization acuity in the azimuth plane. The range of errors and the change in errors with azimuth location shown in Fig. 9 are very similar to the data of Stevens and Newman (1934). Stevens and Newman (1934) used a different procedure than the one used in the current study. If the listeners were able to take advantage of head movements by turning so that the sounds coming from sources to the side were now in front, there would probably not be a clear difference in rms error as a function of relative loudspeaker location as indicated in Fig. 9. So the results of Fig. 9 are indirectly consistent with the assumption that head movements (if they occurred) were not a significant contributing factor for the data collected in this study. Also recall that the rms errors shown in most of the figures of the current paper represent an average of the rms errors across all 11 loudspeaker locations.

(2) Because sounds are processed by cochlear biomechanical mechanisms that act as narrow bandpass filters, there are always envelopes for wideband sounds (i.e., as Rice, 1954, argued a consequence of filtering with a narrow filter is to provide a slow amplitude modulation of the filtered sound). Thus, it is possible that fibers in each auditory nerve that are similarly tuned could provide an envelope ITD cue, when the sound arrives at one ear before the other. It is likely the case that the depth of modulation due to cochlear filtering would be shallow. However, providing any stimulus-based modulation does not appear to have much of an effect on sound source

localization accuracy. It does not seem probable that a stimulus that is modulated with 100% depth of modulation would not affect sound source localization accuracy, but modulation provided by cochlear processing would.

(3) The wideband single click stimulus had a lower rms accuracy error than a short-duration wideband noise stimulus (see inset to Fig. 2). The average rms error of 5.6°–5.7° for a single click stimulus is the smallest rms error we have measured for any stimulus condition tested to date (Yost *et al.*, 2013; Yost and Zhong, 2014; Yost, 2016). In past studies (Yost *et al.*, 2013; Yost and Zhong, 2014; Yost, 2016) using noise stimuli that were two or more octaves wide there appeared to be no difference in rms error as a function of a filter's CF. However, in the current study using clicks, HF clicks that were 2 octaves wide (2–8 kHz) had slightly higher rms errors (1.4° greater rms error) than the BB clicks [125 Hz–8 kHz; see (9) above]. The HF click rms errors (7.9°) are similar to the BB (125 Hz–8 kHz) data shown for the UM noise bursts in Fig. 5 (average rms error of 7.2°). Thus, as for a single click, multiple clicks appear to have slightly lower rms errors than those measured for BB noises. It is not clear why clicks have low rms errors. Perhaps the stochastic aspects of the noise waveforms make it slightly more difficult to localize them than a click whose waveform is not random.

(4) The lead-lag paradigm used to study the effects of precedence for click stimuli (see Litovsky *et al.*, 1999; Brown *et al.*, 2014) involve a two-click (lead and lag) click train often presented at rates of 100 to 1000 Hz (10 to 1 ms interclick interval). In these studies of precedence the first click can clearly dominate the perception of sound source localization (localization dominance, see Litovsky *et al.*, 1999). Thus, localization dominance is clearly different from that obtained in the present study in which the number of clicks had no effect on sound source localization accuracy. However, the fact that the lag click is at a different spatial location than the lead click means the two-click, lead-lag paradigm used to study the effects of precedence is fundamentally different from the click trains used in the present experiment. However, most studies of ITD discrimination thresholds show that ITD thresholds for long duration stimuli (e.g., click trains) are most sensitive to the binaural information at the start of the stimulus (e.g., see Stecker *et al.*, 2013). These results are often modeled with some sort of adaptation process that emphasizes the early part of a waveform consistent with the effects of precedence which indicate the importance of the first arriving sound in processing sound in a reflected space (e.g., see Stecker *et al.*, 2013). The data of the present paper do not indicate much of a change in sound source localization accuracy with changes in the stimuli that occur over time (i.e., modulation). However, as pointed out above, sound source localization accuracy is a different measure than ITD thresholds, especially when ILD is held constant. Additional study is required to determine how ITD threshold and sound source localization accuracy measures are related in terms of envelope variables.

(5) In comparing sound source localization accuracy data across studies the stimuli often vary in overall level and duration (e.g., in Yost, 2016, overall level was 65 dBA and in the present study it was 40 dBA). Because Yost (2016) demonstrated that overall level and duration had little effect on sound source localization accuracy over a considerable range of levels, duration, and filter conditions, it seems unlikely that differences in stimulus level or duration across studies will account for much of the differences in performance that might be apparent.

(6) Several studies (Hartmann *et al.*, 1998; Yost *et al.*, 2013; Yost and Zhong, 2014) have suggested that identification procedures like the one used in the present paper might overestimate sound source localization accuracy because such an identification procedure uses a closed and large spatially-spaced (15° loudspeaker separation) set and a small number of loudspeaker locations that the listener can see. However, this past literature (see Hartmann *et al.*, 1998 for a detail discussion of sound source localization identification procedures) suggests that the 13 loudspeakers spaced 15° apart as used in the procedure of the present paper is sufficient to reduce the effect of these crucial variables in estimating sound source localization accuracy. In addition, the estimates of sound source localization accuracy obtained in this and several other studies (e.g., see Stevens and Newman, 1934; Eberle and Flanagan, 2000; Grantham *et al.*, 2007; Loiselle *et al.*, 2016. as examples) all produce about the same estimates of sound source localization accuracy independent of the type of procedure used to measure sound source localization accuracy. For instance, Stevens and Newman (as the subjects in their study) were blindfolded and indicated the azimuth position of a single loudspeaker at the end of a pole suspended in air in front of them at different azimuth angles. Stevens and Newman's (1934) estimates of sound source localization accuracy were on average 19° for 4 kHz tone, 5.6° for a hiss (noise), and 8° for a click. These compare to 20°, 7.2°, and 5.7° in the current study. So despite very different methods, the estimates of sound source localization accuracy seem to be somewhat similar. There is also no evidence that a difference in sound source localization accuracy measurement procedures would differently affect rms errors as a function of envelope manipulations.

(7) While headphone studies of envelope ITD processing have been extremely useful in understanding the type of binaural processing (e.g., cross correlation) that might be used in sound source localization, this paper along with a few others (Eberle and Flanagan, 2000; Yost and Brown, 2013) suggest that such envelope ITD processing may not provide useful information for judging the location of sounds in an everyday world of listening in a sound field. Eberle and Flanagan (2000) showed that sound field sound source localization for a HF, amplitude-modulated noise was only better than UM noise at a modulation rate of 320 Hz. They argued that this improved sound source localization accuracy was probably a result of a wider bandwidth and not to the

modulation *per se*. Yost and Brown (2013) used amplitude modulation to study sound source localization of two simultaneously presented out-of-phase modulated BB noise bursts. The ability to localize two sound sources was affected by modulation rate (at low modulation rates) when the modulation envelope presented from one source was out-of-phase with that presented from the other source. However, modulation in the Yost and Brown (2013) study had no apparent effect on the ability to localize a single sound source.

(8) The results of the present paper apply to stimuli in the front azimuthal plane and, thus, may not apply to sound source localization accuracy measured for sound sources off this plane. Measurements in the front azimuthal plane were chosen because the studies of envelope ITD processing using headphone delivered stimuli are based on the assumption that ITD cues are important for sound source localization in the front azimuthal plane (i.e., cues other than interaural differences are assumed to play a major role in sound source localization processing off the front azimuthal plane).

## VI. SUMMARY

Studies of envelope ITD processing measured when HF stimuli are delivered over headphones have shown that envelope ITDs can affect ITD discrimination thresholds and laterality judgments. Such results have been and continue to be important in revealing properties of binaural processing that might be crucial for sound source localization in the front azimuth sound field where interaural cues are the crucial ones for sound source localization. However, the results of the present studies and the few others (Eberle and Flanagan, 2000; Yost and Brown, 2013) that have investigated envelope cues in sound source localization in a sound field have shown that envelope ITD cues play either no role or a very small role in sound source localization accuracy in the front azimuth sound field.


## ACKNOWLEDGMENT

The research was supported by a grant from the Air Force Office of Scientific Research (AFOSR) and a grant from the National Institute on Deafness and Other Communication Disorders (NIDCD) both awarded to W.A.Y. The author is grateful to Dr. Michael Dorman, Dr. Yi Zhou, and Dr. Torben Pastore for comments on this research project. The assistance of Kathryn Pulling and Anbar Najam is also gratefully acknowledged.


[1]In the Yost (2016) paper sound source localization accuracy was measured as a function of noise duration for different noise filter conditions, including a BB and a narrowband condition exactly like the filters used in the present experiment. A single click filtered the same as the noises was also used in Yost (2016). Only the noise data were reported in the Yost (2016) paper. The current paper includes data comparing the single filtered click data to the 25 ms duration filtered noise data of Yost (2016). Twelve listeners were used in the Yost (2016) paper based on the Yost *et al.* (2013) paper which used a large number (45) of listeners to suggest that 12 listeners produced sufficient statistical power in measuring sound source localization accuracy. Twelve listeners are used in all of the experiments of this paper for the same reason.

[2]The first click in the click trains started at time zero and the next clicks started at time intervals equal to the reciprocal of the click rate (when expressed in Hz). The audio files containing the click trains were longer than the time from first click onset to last click onset to allow for the ringing of the clicks. The file length was set to the length between the first click onset to the last click onset plus 15.625 ms.


Bernstein, L. R., and Trahiotis, C. (**1985**). "Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities," J. Acoust. Soc. Am. **77**, 1868–1880.

Bernstein, L. R., and Trahiotis, C. (**1994**). "Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes and bands of noise," J. Acoust. Soc. Am. **95**, 3561–3567.

Bernstein, L. R., and Trahiotis, C. (**2002**). "Enhancing sensitivity to interaural delays at high frequencies by using 'transposed stimuli,'" J. Acoust. Soc. Am. **112**, 1026–1036.

Bernstein, L. R., and Trahiotis, C. (**2009**). "How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli," J. Acoust. Soc. Am. **125**, 3234–3242.

Bernstein, L. R., and Trahiotis, C. (**2014**). "Sensitivity to envelope-based interaural delays at high frequencies: Center frequency affects the envelope rate-limitation," J. Acoust. Soc. Am. **135**, 808–816.

Bibee, J. M., and Stecker, G. C. (**2016**). "Spectrotemporal weighting of binaural cues: Effects of a diotic interferer on discrimination of dynamic interaural differences," J. Acoust. Soc. Am. **140**, 2584–2592.

Blauert, J. (**1997**). *Spatial Hearing* (The MIT Press, Cambridge, MA), 494 pp.

Brown, A. D., Stecker, G. C., and Tollin, D. J. (**2014**). "The precedence effect in sound localization," J. Assoc. Res. Otolaryngol. **18**, 1–28.

Buell, T. N., Trahiotis, C., and Bernstein, L. R. (**1994**). "Lateralization of bands of noise as a function of combinations of interaural intensive differences, interaural temporal differences, and bandwidth," J. Acoust. Soc. Am. **95**, 1482–1489.

Diedesch, A. C., and Stecker, G. C. (**2015**). "Temporal weighting of binaural information at low frequencies: Discrimination of dynamic interaural time and level differences," J. Acoust. Soc. Am. **138**, 125–133.

Eberle, G., and Flanagan, P. (**2000**). "Localization of amplitude-modulated high-frequency noise," J. Acoust. Soc. Am. **107**, 3568–3571.

Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Labadie, R. F., and Haynes, D. S. (**2007**). "Horizontal-plane localization of noise and speech signals by postlingually deafened adults fitted with bilateral cochlear implants," Ear Hear. **28**, 524–541.

Hafter, E. R., Dye, R. H., and Wenzel, E. (**1983**). "Detection of interaural differences of intensity in trains of high-frequency clicks as a function of interclick interval and number," J. Acoust. Soc. Am. **73**, 1708–1713.

Hartmann, W. M., Rakerd, B., and Gaalaas, J. B. (**1998**). "On the source-identification method," J. Acoust. Soc. Am. **104**, 3546–3557.

Houtgast, T., and Steeneken, H. J. M. (**1973**). "The modulation transfer function in room acoustics as a predictor of speech intelligibility," Acustica **28**, 66–73.

Kuhn, G. F. (**1987**). "Physical acoustics and measurements pertaining to directional hearing," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York), pp. 3–25.

Litovsky, R. Y., Colburn, H., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Loiselle, L. H., Dorman, M. F., Yost, W. A., Cook, S. J., and Gifford, R. H. (**2016**). "Using ILD and ITD cues for sound source localization and speech understanding in complex listening environment by listeners with bilateral and with hearing-preservation cochlear-implants," J. Speech Lang. Hear. Res. **59**, 810–818.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246.

Monaghan, J. J. M., Bleeck, S., and McAlpine, D. (**2016**). "Sensitivity to envelope interaural time differences at high modulation rates," Trend. Hear. **19**, 1–14.

Rakerd, B., and Hartmann, W. M. (**1986**). "Localization of sound in rooms III. Onset and duration effects," J. Acoust. Soc. Am. **80**, 1695–1706.

Rakerd, B., and Hartmann, W. M. (**2010**). "Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences," J. Acoust. Soc. Am. **128**, 3052–3063.

Rayleigh, Lord. (J. W. Strutt) (**1907**). "On our perception of sound direction," in Philosophical Magazine, 136th series, pp. 456–464.

Rice, S. O. (**1954**). "Mathematical analysis of random noise," in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York).

Stecker, G. C., and Bibee, J. M. (**2014**). "Nonuniform temporal weighting of interaural time differences in 500 Hz tones," J. Acoust. Soc. Am. **135**(6), 3541–3547.

Stecker, G. C., Ostreicher, J. D., and Brown, A. D. (**2013**). "Temporal weighting functions for interaural time and level differences. III. Temporal weighting for lateral position judgments," J. Acoust. Soc. Am. **134**, 1242–1252.

Stevens, S. S., and Newman, E. B. (**1934**). "The localization of actual sources of sound," Am. J. Psych. **48**, 297–306.

Trahiotis, C., and Stern, R. M. (**1989**). "Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase," J. Acoust. Soc. Am. **86**, 1285–1293.

Yost, W. A. (**1976**). "Lateralization of repeated filtered transients," J. Acoust. Soc. Am. **60**, 178–181.

Yost, W. A. (**2007**). *Fundamentals of Hearing: An Introduction* (Academic Press, New York), 336 pp.

Yost, W. A. (**2016**). "Sound source localization identification accuracy: Level and duration dependencies," J. Acoust. Soc. Am. **140**, 14–19.

Yost, W. A., and Brown, C. A. (**2013**). "Localizing the sources of two independent noises: Role of time varying amplitude differences," J. Acoust. Soc. Am. **133**, 2301–2313.

Yost, W. A., and Hafter, E. R. (**1987**). "Lateralization," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York), pp. 49–84.

Yost, W. A., Loiselle, L., Dorman, M., Brown, C., and Burns, J. (**2013**). "Sound source localization of filtered noises by listeners with normal hearing: A statistical analysis," J. Acoust. Soc. Am. **133**, 2876–2882.

Yost, W. A., Wightman, F. L., and Green, D. M. (**1971**). "Lateralization of filtered clicks," J. Acoust. Soc. Am. **50**, 1526–1531.

Yost, W. A., and Zhong, X. (**2014**). "Sound source localization identification accuracy: Bandwidth dependencies," J. Acoust. Soc. Am. **136**, 2737–2746.

Yost, W. A., Zhong, X., and Najam, A. (**2015**). "Rotating sound sources and listeners: Sound source localization is a multisensory/cognitive process," J. Acoust. Soc. Am. **138**, 3293–3310.